PERCEPTION OF INTONATIONAL CHARACTERISTICS OF WH AND NON-WH QUESTIONS IN TOKYO JAPANESE

Kikuo Maekawa
National Language Research Institute, Tokyo, Japan.

ABSTRACT
The intonational difference between wh and non-wh questions in Tokyo Japanese was examined. Perception experiments involving synthetic intonation revealed that the most important cue for the discrimination between the two types is the lack of prominence of the focused wh-word, rather than the prominence of the focused wh-word per se.

1. INTRODUCTION
That syntactic behavior of wh and non-wh questions differs is well recognized by grammarians. It seems to be less recognized by those who are working with Japanese prosody that the two question types differ significantly in their prosodic domains as well. As a matter of fact, the difference does not consist in a mere difference of final rise but rather concerns the overall intonation shapes.

2. MATERIAL
Wh-questions are marked with wh-words like dare (who), doko (where), nanik (what) etc. Incidentally, there are a class of words which are not wh-words but morphologically very similar to them: dareka (someone), dekoka (somewhere), nanika (something) etc. Those words are semantically marked, given their indefinite-pronoun-like meaning. As the result of that syntactic similiarity, we can construct

3. EXPERIMENT
The aim of the first experiment was to examine if native speakers of TJ can in fact discriminate the two question types solely by means of intonation. The difference of (1) and (2) consists in the /k/-/g/ consonantal contrast as far as the segmental tier is concerned. So it was expected that subjects would be forced to rely on prosodic cues if we erased these consonants and then filled the resulting silence with white noise. In this reasoning, the following ten stimuli were prepared. The underlines show the time stretch replaced with noise.

4. RESULTS
In erasing sequences of segments, care was taken to rid the effect of coarticulation as much as possible. Consequently, the white noise penetrates more or less into the final part of preceding segment and the beginning of following segment in all cases. All manipulation of original utterances, which were sampled in 10kHz/16bits
The aim of the second experiment was to examine the program developed by Hiroshi Mizutani of NLRI for his contribution.) The remaining 14 points are represented by the same listeners as in the previous experiment. Fig. 4 shows the percentages with which each stimulus was perceived as a wh-question. The abscissa of the figure is a composite representation of D values for the stimulus with E=300Hz (the leftward three values) and for the stimulus with E=800Hz (the rest). The real and dotted lines stand respectively for the stimuli with B=300Hz and B=230Hz. This figure shows clearly that the contribution of the D value is greater by far than that of the B value. Although a raised B value (300Hz) makes some contribution to subjects' judgment of wh-questions, this effect is observed only when D is relatively high (180Hz or 160Hz). Once D is set to relatively low (120Hz or 100Hz), the stimulus was perceived mostly as wh-question irrespective of the B value.

5. DISCUSSION AND CONCLUSION

The two experiments reported here lead us to reconsider the phonetic nature of focus in TJ, stressing the importance of the salience of the prosodic boundary. In this respect, Fujisaki & Kawai (1990) point out that focus not only increases the prominence of the focused constituent but also reduces the prominence of the following constituents. Kori (1987) has independently pointed out that focus may not necessarily reduce the prominence of the following constituents. Kori also suggests that prominence of the final constituent of an utterance is more reduced than that of the other constituents. This analysis, which is based on production data, seems to be congruent with my perception data. Fig. 4 indicates that in order for a stimulus to be identified as a wh-question with 90% accuracy, it is necessary that the B value be lower than 120Hz i.e. lower than the right edge of the preceding NP. The data presented here and that of Kori and that of Fujisaki & Kawai suggest that any theory of phonetics that assumes that the effect of focus is limited only to the constituent marked as focused is inappropriate and to be revised. The two experiments reported here revealed that one important problem was left untouched: whether the difference of intonation examined in this study is specific to the pair of wh and non-wh questions. The line of reasoning that I followed in this study predicts that the difference is not a specific one. It is expected that the same intonational difference is observed in any pair of sentences having the same difference of focus placement as the one observed between (3) and (4).

6. REFERENCES


ACKNOWLEDGMENT

I am grateful to Osamu Mizutani of NLRI for his comments on an earlier draft of this paper.

300
200
100
HZ
WN
Fig. 3: Schematic structure of the synthetic stimuli. Control points A-F were linearly interpolated as a gross approximation to natural intonations. The thick arrow indicates the time stretch masked with white noise.

Fig. 4: wh-judgments of sixteen synthetic stimuli as the function of the D values (abscissa). Real lines stand for the stimulus with B=300Hz (prominent wh), and dotted lines for those with B=230Hz (not prominent).